

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-083399

(43)Date of publication of application : 31.03.1998

(51)Int.Cl.

G06F 17/27

G06F 17/28

(21)Application number : 09-125213

(71)Applicant : SHARP CORP

(22)Date of filing : 15.05.1997

(72)Inventor : KUTSUMI TAKESHI

(30)Priority

Priority number : 08122049

Priority date : 16.05.1996

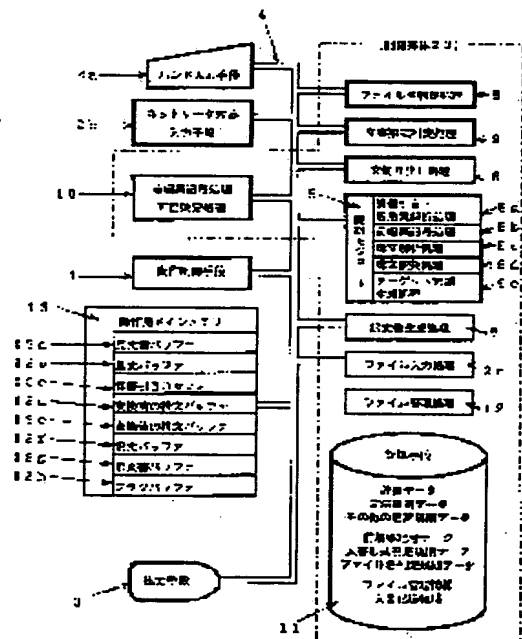
Priority country : JP

(54) MACHINE TRANSLATION DEVICE AND COMPUTER-READABLE RECORDING MEDIUM WHERE TRANSLATING PROGRAM IS RECORDED

(57)Abstract:

PROBLEM TO BE SOLVED: To save user's trouble and simplify operation by deciding whether a translation document is created according to editing symbols or not.

SOLUTION: A translation module 5 creates the translation document. A segmentation processing means 6 cuts one sentence out of the document. A translation document generation processing means 7 reconstitutes a translation document from a translation obtained in sentence units. A file name decision processing means 8 decides whether or not a file name meets specific conditions when reading a document out of a file. A document format decision processing means 9 decides whether or not the format of the contents of the document meets specific conditions. Then an editing symbol processable/unprocessable state decision means 10 determines whether or not editing symbols are processed according to the result of the condition decision making. Thus, it is decided whether or not document format decision rule data are met to determine whether the translation document is generated according to the editing symbols or not according to the decision result.

**LEGAL STATUS**

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-83399

(43) 公開日 平成10年(1998) 3月31日

(51) Int.Cl.⁶
G 0 6 F 17/27
17/28

識別記号 庁内整理番号

F I
G 0 6 F 15/38

技術表示箇所

D
Q

審査請求 未請求 請求項の数6 O L (全 15 頁)

(21) 出願番号 特願平9-125213

(22) 出願日 平成9年(1997) 5月15日

(31) 優先権主張番号 特願平8-122049

(32) 優先日 平8 (1996) 5月16日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 九津見 毅

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

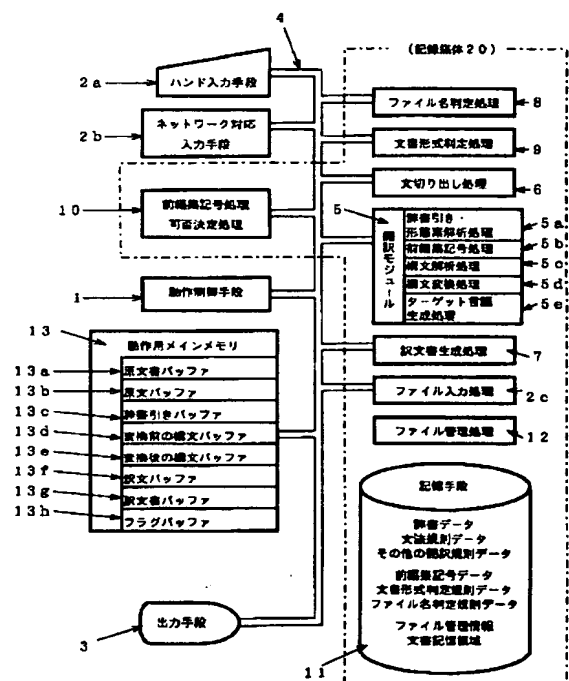
(74) 代理人 弁理士 梅田 勝

(54) 【発明の名称】 機械翻訳装置及び翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 前編集記号と誤認する恐れのある記号を有する文書の機械翻訳において、利用者が前編集記号の有効・無効の設定を行う手間を省いて操作性を良好にする。

【解決手段】 機械翻訳装置や翻訳プログラムを記録した媒体が、文書内容の形式、文書ファイル名、使用した文書入力手段の種類等の各種状況を検出し、その結果に応じて、翻訳の際に前編集記号処理を行うか否かを選択する。



【特許請求の範囲】

【請求項 1】ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、

文書形式判定規則データを記憶している記憶手段と、上記原文書の文書形式が上記文書形式判定規則データに合致するか否かを判定する文書形式判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段とを備えていることを特徴とする機械翻訳装置。

【請求項 2】コンピュータにおいて、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体であって、前記翻訳プログラムは、記憶手段に文書形式判定規則データを記憶させ、制御手段に、原文書の文書形式が文書形式判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させることを特徴とする翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 3】ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、ファイル名判定規則データを記憶している記憶手段と、上記原文書のファイル名が上記ファイル名判定規則データに合致するか否かを判定するファイル名判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段とを備えていることを特徴とする機械翻訳装置。

【請求項 4】コンピュータにおいて、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体であって、前記翻訳プログラムは、記憶手段にファイル名判定規則データを記憶させ、制御手段に、原文書の文書形式がファイル名判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成する

か、前編集記号に従わずに訳文書を生成するかを決定させることを特徴とする翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 5】入力手段から入力されるソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、

10 上記入力手段の種類如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段を備えていることを特徴とする機械翻訳装置。

【請求項 6】コンピュータにおいて、入力手段より入力されるソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体であって、前記翻訳プログラムは、

20 制御手段に、入力手段の種類を判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させることを特徴とする翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、翻訳の際に利用者が予め原文に前編集処理を施すことによって翻訳効率や精度を高めることができる機械翻訳処理に関するものである。

【0002】

【従来の技術】一般的な機械翻訳装置としては次のようなものがある。図 20 に示すように、この機械翻訳装置は、キーボード 104 等の入力手段から入力されたソース言語を、CPU 101 やメインメモリ 102 等の制御手段の制御により翻訳モジュール 105 に入力し、この入力されたソース言語を、メモリ 106 等の記憶手段に記憶されている辞書、文法規則および木構造変換規則を利用して、翻訳モジュール 105 によってターゲット言語に翻訳し、表示部 103 等の上に表示するものである。

【0003】このような機械翻訳装置では、形態素解析の結果得られた品詞から、辞書と文法規則を用いて構文解析を行うとき、あらゆる構文構造を組み立てるようにしており、複数の翻訳結果を出力していた。そのため、複数の翻訳結果のうちどれが正しい翻訳結果であるかを利用者が選択する必要があった。また、あらゆる構文構造を組み立てるために構文解析に長い時間がかかっていた。

【0004】この問題を少しでも減らすため、翻訳を実

行する前にあらかじめ利用者が原文に前編集処理を施すことにより、構文構造の候補を限定したり、構文解析の速度を速めたりすることができる機械翻訳装置が今までに考案されてきた。例えば特開平1-113870号公報に開示された技術がある。ここでいう前編集処理とは、通常の文章には用いられないような特殊な文字・記号列を何種類か「前編集記号」として定義し、それを原文の文中に挿入することにより、特定の単語の品詞を指定したり、構文構造を指定したりすることである。

【0005】また、そのような前編集記号が、文に使われている文字列や記号と一致することがあり、その場合は当該箇所は前編集記号であるかのように機械翻訳装置によって誤認されてしまい、正しい翻訳が行われないう問題があった。しかし、この問題に対しては、前編集記号の形を利用者が自由に定義することができるような機械翻訳装置を用いることによって解決することができる。このような機械翻訳装置は特開平5-334347号公報に示されている。それは例えば前述の図20に示した構成を持ち、また、図21に示すように、利用者によって定義された前編集記号データが格納されたバッファを有している。

【0006】

【発明が解決しようとする課題】上述したように、従来は、前編集記号を用いることにより翻訳効率の向上を図るような機械翻訳装置がある。また、前編集記号が文書に使われている文字列や記号と一致することがあるという問題に対しては、前編集記号の形を利用者が自由に定義する機械翻訳装置によって解決を図ることができる。しかし、このような機械翻訳装置には、そのようにして前編集記号を変更する操作それ自体が煩わしいという問題点がある。

【0007】近年、通信ネットワークの発達に伴い、文書の体裁情報を記述したマークアップ・タグ記号や、ネットワーク上のアドレス情報などの付いた文書を機械翻訳する機会が増えてきている。このような文書を翻訳しようすると、前編集記号の形によっては、文書中の文字列が前編集記号と偶然に一致する可能性が高い場合がある。

【0008】たとえば、前編集記号が図3のように定められている機械翻訳装置を用いて、図7に示すような文書を翻訳した場合を考える。図3に示すように、「／／」という文字列からなる前編集記号が文書中にあると、そこで文が分割されることになっている。ところが、この文書中には、ネットワーク・アドレス情報を示す文字列「http：／／www.sharp.co.jp／」が存在する。前編集記号を変更する操作を事前に行わないで翻訳すると、上記文字列中の「／／」の箇所が、文分割を指示する前編集記号であると解釈される。そのため、翻訳結果は、図15に示すように、ひとまとまりであるべきアドレス情報「http：／／ww

w.sharp.co.jp／」が分割されてしまう。

【0009】このため、上述のような機械翻訳装置を用いてこのような文書を翻訳しようとする、必ず、前編集記号を変更する操作を事前に行っておく必要がある。すなわち、この煩わしい操作を必ず行わなければならない。

【0010】本発明は上記問題点に鑑みなされたものであり、その目的は、前編集記号を変更する操作を装置の利用者がわざわざ行わなくても、翻訳を実行する際の状況に応じて前編集記号の扱いを全く無効とするように翻訳動作を自動的に変更することができ、それによって、操作を簡略化することができる機械翻訳装置及び翻訳プログラムを記録した媒体を提供することにある。

【0011】

【課題を解決するための手段】上記の課題を解決するために、請求項1記載の機械翻訳装置は、ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、文書形式判定規則データを記憶している記憶手段と、上記原文書の文書形式が上記文書形式判定規則データに合致するか否かを判定する文書形式判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段とを備えていることを特徴としており、請求項2記載の翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体は、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、記憶手段に文書形式判定規則データを記憶させ、制御手段に、原文書の文書形式が文書形式判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させることを特徴としている。

【0012】上記請求項1乃至2記載の構成により、ソース言語の文書が入力された際に、文書の形式が判定される。この結果、前編集記号処理を無効にすべき文書かどうか判定される。この判定のために、記憶手段に記憶された文書形式判定規則データが参照される。

【0013】原文書が前編集記号処理が無効である文書であると判定された場合は、前編集記号処理の可否決定を行い、前編集記号処理を行わずに翻訳される。

【0014】一方、原文書が前編集記号処理が有効な文書であると判定された場合は、前編集記号処理の可否決定を行うことにより、前編集記号によって指定された条件に沿った形で翻訳される。また、前編集記号によって指定されている条件によっては、原文書から1文を切り

出す1文切り出し処理の修正が行われる。このようにして、前編集記号による指定条件を有効にした翻訳が実行される。

【0015】すなわち、文書形式の判定によって、前編集記号と誤認する可能性のある記号類を含む形式の文書であるかどうか判定され、その結果に応じて前編集記号処理の可否決定を行い前編集処理が無効化される。

【0016】したがって、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が翻訳実行に先立ちあらかじめ決定しておく必要がない。つまり、あらかじめ利用者の手で前編集記号を初期設定から変更あるいは無効化しておく必要がない。それによって、文書の翻訳において、利用者の手間を省き、操作を簡略化することができる。

【0017】また、請求項3記載の機械翻訳装置は、ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、ファイル名判定規則データを記憶している記憶手段と、上記原文書のファイル名が上記ファイル名判定規則データに合致するか否かを判定するファイル名判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段とを備えていることを特徴としており、請求項4記載の翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体は、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、記憶手段にファイル名判定規則データを記憶させ、制御手段に、原文書の文書形式がファイル名判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させることを特徴としている。

【0018】上記請求項3乃至4記載の構成により、ソース言語の文書が入力された際に、ファイル名が判定される。この結果、前編集記号処理を無効にすべき文書かどうか判定される。この判定のために、記憶手段に記憶されたファイル名判定規則データが参照される。

【0019】つまり、ファイル名の判定によって、前編集記号と誤認する可能性のある記号類を含むような傾向を持つ文書であるかどうか判定され、その結果に応じて前編集記号処理可否決定手段によって前編集処理が無効化される。

【0020】したがって、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が決定する必要がない。つまり、あらかじめ利用者の手で前

編集記号を初期設定から変更あるいは無効化しておく必要がない。それによって、文書の翻訳において、利用者の手間を省き、操作を簡略化することができる。

【0021】また、請求項5記載の機械翻訳装置は、入力手段から入力されるソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、上記入力手段の種類如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段を備えていることを特徴としており、請求項6記載の翻訳プログラムを記録したコンピュータ読み取り可能な記録媒体は、入力手段より入力されるソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、制御手段に、入力手段の種類を判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させることを特徴としている。

【0022】上記請求項5乃至6記載の構成により、ソース言語の文書が入力される際に、その入力手段の種類如何によって、前編集記号処理を無効にすべきか否かが判定される。

【0023】つまり、原文書の入力の際に使用された入力手段の種類に基づき、前編集記号と誤認する可能性のある記号類を含むような傾向を持つ文書であるかどうか判定され、その結果に応じて前編集記号処理の可否決定を行い、前編集処理が無効化される。

【0024】したがって、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が決定する必要がない。つまり、あらかじめ利用者の手で前編集記号を初期設定から変更あるいは無効化しておく必要がない。それによって、文書の翻訳において、利用者の手間を省き、操作を簡略化することができる。

【0025】

【発明の実施の形態】

〔実施の形態1〕本発明の実施の一形態について図1ないし図15に基づいて説明すれば、以下の通りである。本実施の形態では、ソース言語が英語であり、ターゲット言語が日本語である場合を例にとり説明するが、これらの言語は特にこれらに限る必要はない。

【0026】図1は本発明の機械翻訳装置や翻訳プログラムを記録した媒体における一実施の形態の構成を示す図である。動作制御手段1、ハンド入力手段2a、ネットワーク対応入力手段2b、ファイル入力処理装置／プログラム2c、出力手段3、翻訳モジュール5、文切り出し処理手段6、訳文書生成処理手段7、ファイル名判

定処理手段8、文書形式判定処理手段9、前編集記号処理可否決定処理手段10、記憶手段11、ファイル管理処理手段12、および動作メインメモリ13が、バスライン4を介して相互に接続され、各プログラムの実行が行われる。

【0027】動作制御手段1は、実際の機器構成においては主としてCPU（中央処理装置）をもってこれを実現する。

【0028】ハンド入力手段2aは装置の利用者が手入力するためのものであり、実際の機器構成においてはキーボード・マウス・ペン・タブレット・スキャナ・文字認識装置などの入力装置をもってこれを実現する。ネットワーク対応入力手段2bは、通信回線と接続されている通信装置をもってこれを実現する。ファイル入力処理手段2cは、該機械翻訳装置や翻訳プログラムを記録した媒体によって記憶手段11に記憶されている文書情報を読み出すためのものである。

【0029】出力手段3は、実際の機器構成においてはCRT（陰極線管）・LCD（液晶表示装置）などの表示装置や、プリンターなどの印字装置をもってこれを実現する。

【0030】翻訳モジュール5は後述の動作により翻訳文書を作成するものである。文切り出し処理手段6は文書から1文を切り出すためのものである。訳文書生成処理手段7は1文単位で訳された訳文から翻訳文書を再構成するものである。ファイル名判定処理手段8はファイルから文書を読み込んだ場合にファイル名が所定の条件に合致するかを判定するものである。文書形式判定処理手段9は文書の中身の形式が所定の条件に合致するかを判定するものである。前編集記号処理可否決定処理手段10は条件判定の結果として前編集記号の処理を行うか否かを決定するものである。

【0031】実際の機器構成においては、上記ファイル入力処理手段2c、翻訳モジュール5、文切り出し処理手段6、訳文書生成処理手段7、ファイル名判定処理手段8、文書形式判定処理手段9および前編集記号処理可否決定処理手段10はいずれも、CPU（中央処理装置）とコンピュータプログラムとによって構成される機能モジュールとすることができる。上記CPUとして前記した動作制御手段1が用いられ、また、各コンピュータプログラムは、記録媒体20に格納されて、実行時は動作メインメモリ13に読み込まれ、実行されることにより各手段として動作するようにすることができる。

【0032】記憶手段11は、翻訳用の辞書・文法規則その他翻訳に必要な規則等のデータ、および、前編集処理に必要な規則や前編集処理の可否決定に必要な規則などのデータを格納しているものである。ファイル管理処理手段12は記憶手段11をファイルシステムとして使用するためのものであり、これによって記憶手段11には、翻訳するための文書そのものも記憶することができ

る。動作メインメモリ13は、翻訳作業中に必要な各種データを記憶するためのものであり、内部にそのための各種バッファを備えている。

【0033】実際の機器構成においては、上記記憶手段11および動作メインメモリ13は、それぞれハードディスク等やRAM（Random Access Memory）、前記録媒体20などで構成し、また、上記ファイル管理処理手段12は、CPUとコンピュータプログラムとによって構成される機能モジュールとすることができる。上記CPUとして前記した動作制御手段1が用いられ、また、上記コンピュータプログラムは、記録媒体20に格納されて、実行時は動作メインメモリ13に読み込まれるようにすることができる。

【0034】一般に、機械翻訳には図2に示すような解析レベルがある。図中左上においてソース言語が入力されると、レベルL1の辞書引き、レベルL2の形態素解析、レベルL3の構文解析、…と、解析が進められる。

【0035】機械翻訳はこの解析レベルにより、大きく次の2つにわけられる。1つは、レベルL6のソース言語およびターゲット言語のどちらにも依存しない概念（中間言語と呼ぶ）まで解析し、そこから、レベルL7の文脈生成、レベルL8の意味生成、レベルL9の構文生成、レベルL10の形態素生成と生成を進めて、ターゲット言語を生成していくピボット方式である。

【0036】もう1つは、上述のレベルL2の形態素解析、レベルL3の構文解析、レベルL4の意味解析およびレベルL5の文脈解析のいずれかまで解析を行ってソース言語の内部構造を得、次に、この得られたソース言語の内部構造と同レベルのターゲット言語の内部構造に変換した後、ターゲット言語を生成するトランスファー方式である。

【0037】ここで、上記各解析について説明する。

【0038】①辞書引き、形態素解析

図1に示す記憶手段11の辞書データを用いて辞書を引き、入力された文章を各形態素列（単語列）に分割し、この各単語に対する品詞などの文法情報および訳語を得、さらに、時制・人称・数などを解析する。

【0039】②構文解析

単語間の係り受けなどの、文章の構造（構造解析木）を決定する。

【0040】③意味解析

複数の構文解析の結果から意味的に正しいものとそうでないものを判別する。

【0041】④文脈解析

話題を理解し、省略や曖昧さを取り去る。

【0042】本実施の形態で用いる翻訳モジュール5は、少なくとも図2で説明したレベルL3の構文解析のレベルまでの解析を行うものとする。

【0043】この原理による翻訳を実現するために、翻訳モジュール5の構成要素には、図1に示すように、辞

10

20

30

40

50

書引き・形態素解析処理手段5a、前編集記号処理手段5b、構文解析処理手段5c、構文変換処理手段5d、ターゲット言語生成処理手段5eを備えている。前編集記号処理手段5bは、前編集記号を解釈して処理するためのものである。前編集記号は、図3に示すように定められていて、記憶手段11に記憶されている。

【0044】また、動作メインメモリ13は、翻訳動作の進行に応じて、その一部が、原文書バッファ13a、原文書バッファ13b、辞書引きバッファ13c、変換前の構文バッファ13d、変換後の構文バッファ13e、訳文バッファ13f、訳文書バッファ13g、フラグバッファ13hに、それぞれ割り当てられる。

【0045】本実施の形態の場合、文書形式判定のための規則が図4に示すように定められていて、記憶手段11に記憶されている。ここでは、2種類の条件を図示している。すなわち、1つ目は、文書中に「http://」という文字列が含まれているという条件である。2つ目は、文書内容が「<html>」で始まり「</html>」で終わっているという条件である。これらの条件のうちのどれか一つが適合すれば、前編集記号は無効にされる。図中、IDとは、前編集記号処理を行わない条件の番号を示すものである。

【0046】本実施の形態による装置の動作の過程は、図5と図6のフローチャートに示すようになっている。まず図5に示すように動作し、図5に示す過程を終了したら図6に示すように動作する。

【0047】ここでは、図7に示すような文書を翻訳するものとして動作を説明する。まず、図5のステップS1で、フラグ変数peのバッファ領域がフラグバッファ13hに確保され、フラグ変数peに値0が格納される。次にS2で、図7の原文書がハンド入力手段2aから読み込まれ、原文書バッファ13aに格納される。次にS3で、文書形式判定処理手段9によって、原文書バッファ13aに格納されている原文書の全文が検索される。この検索は原文書がネットワーク上の文書としての特徴を持ったものかを決定するためのもので、この決定のための規則は記憶手段11に記憶されている。次に、S4で、原文書が、前編集記号処理を有効にすべき文書か無効にすべき文書かを判定する。すなわち、ここでは、この規則は図4に示すようなものであるとする。今の場合、図7に示した原文書は、文字列「<html>」で始まって「</html>」で終わっている。また、文字列「http://」を含んでいる。このような点が文書形式判定処理手段9によって検出され、結果、この文書はネットワーク上の文書としての特徴を持ったもの、すなわち前編集記号処理を行わないものと決定され、S4から1000に進む。

【0048】もし、原文書バッファ13aに格納されている原文書はネットワーク上の文書としての特徴を持たないものであると文書形式判定処理手段9によって判定

されたら、S4からS5に進み、フラグ変数peに値1が格納され、1000に進む。

【0049】以下、図6のフローチャートに沿って説明する。S101で、アドレス変数ptのバッファ領域がフラグバッファ13hに確保され、アドレス変数ptには原文書バッファ13aにおける先頭位置アドレスが格納される。次にS102で、原文書バッファ13aに格納されている原文書から、文切り出し処理手段6によって1文が切り出されて原文書バッファ13bに格納される。このとき切り出される1文は、原文書バッファ13aにおいて、アドレス変数ptに格納されているアドレス値以降の最初に出現する1文である。

【0050】今の場合、アドレス変数が原文書バッファ13aの最初の位置にある。この位置以降で、最初に出現し、文切り出し処理手段6によって文として認定される文字列は、たとえば「<html>」である。

【0051】ここで1文が切り出されると、原文書バッファ13aにおいていま切り出された1文の直後のアドレス値が、アドレス変数ptに格納される(S103)。今の場合、アドレス変数ptは原文書バッファ13aにおける「<html>」の直後の位置を示すアドレス値である。

【0052】次にS104で、前編集記号処理を行うか否かの処理の分岐を行う。今の場合、フラグ変数peが0なので、前編集記号処理を行わないことが示されており、よってS105に進む。

【0053】以下、通常の機械翻訳の過程である辞書引き・形態素解析、構文解析、構文変換、ターゲット言語生成が行われて、1文の翻訳が完了する(S105、S106、S113、S114)。今の場合、原文「<html>」に対する訳文は「<html>」となり、これが訳文バッファ13fに格納される。次にS115で、訳文書生成処理手段7によって、訳文バッファ13fに格納されている訳文が訳文書バッファ13gに追加的に格納される。

【0054】次にS116で、文書の末尾まで訳されたかどうか判定される。今の場合、アドレス変数ptは原文書バッファ13aにおいて原文書の「<html>」の直後の位置を示していて、末尾位置アドレスに達していないので、まだ文書の末尾まで訳されていないとされ、S102に戻る。

【0055】以下、S102からを繰り返す。すなわち、次に文切り出し処理手段6によって切り出される1文は、たとえば今の場合「<body>」である。そしてS103、S104、S105、S106、S113、S114と進んで、訳文「<body>」が得られ、訳文バッファ13fに格納される。そして、S115で訳文が訳文書バッファ13gに追加的に格納される。

【0056】再びS102からを繰り返す。次に文切り

出し処理手段6によって切り出される1文は、「The URL is http://www.sharp.co.jp/」である。S102でこの1文が図8のように原文バッファ13bに格納され、S103で原文書バッファ13aにおける「The URL is http://www.sharp.co.jp/」の直後の位置を示すアドレス値がアドレス変数ptに格納される。

【0057】次にS104で、フラグ変数peが0なのでS105に進む。以下同様に、辞書引き・形態素解析、構文解析、構文変換、ターゲット言語生成が行われて、1文の翻訳が完了する(S105、S106、S113、S114)。今の場合、原文「The URL is http://www.sharp.co.jp/」に対する訳文は「URLは、http://www.sharp.co.jp/である。」となり、これが図9のように訳文バッファ13fに格納される。

【0058】もし、何らかの理由で、フラグ変数peが1、すなわち前編集記号処理が行われるとされていると、S104からS107に進み、辞書引き・形態素解析の前に、前編集記号処理手段5bによって前編集記号処理が行われる。このとき、記憶手段11に記憶されている図3のような形・機能の前編集記号データが参照される。今、原文バッファ13bの内容が図8のように「The URL is http://www.sharp.co.jp/」であるとする、図3の前編集記号データと照合した結果、文分割を指示する前編集記号「／／」が文中に含まれている。よってS108では文が分割されていると判定され、S109に進む。S109では分割位置以降を原文バッファ13bから消し、この結果原文バッファ13bは図10のようになる。そしてS110ではアドレス変数ptの修正が行われ、今の場合は原文書バッファ13aにおける「The URL is http://」の直後の位置を示すアドレス値がアドレス変数ptに格納される。

【0059】S111で辞書引き・形態素解析が行われ、S112で前編集規則に従って構文解析が行われる。構文変換とターゲット言語生成は、前編集記号処理しない場合と同様にステップS113、S114で行われる。この結果、訳文は「URLは、http:である」となり、S115で図11のように訳文バッファ13fに格納され、この訳文が訳文書バッファ13gに追加的に格納される。

【0060】そしてS116からS102に戻り、次の1文が「www.sharp.co.jp/」のように切り出され、図12のように原文バッファ13bに格納される。以下、S104、S107、S111、S112、S113、S114と翻訳処理が進んで、この結果、訳文は「www.sharp.co.jp/。」となり、S115で図13のように訳文バッファ13fに

格納され、S116でこの訳文が訳文書バッファ13gに追加的に格納される。

【0061】以上で、フラグ変数peが1である場合の説明を終わり、フラグ変数peが0である場合の説明に戻る。

【0062】原文書バッファ13aの「The URL is http://www.sharp.co.jp/」より後の部分も同様に1文ずつ翻訳処理が行われ、訳文が訳文書バッファ13gに追加され、最終的にアドレス変数ptの値が原文書バッファ13aの末尾位置アドレスの値となった時点で、S116からendに進んで、翻訳処理が終了する。この結果、訳文書バッファ13gの内容は図14に示すような翻訳文書となる。

【0063】一方、もし、フラグ変数peが1であると、最終的に、訳文書バッファ13gの内容は図15に示すようになる。

【0064】〔実施の形態2〕本発明の他の実施の形態について図1、図6、図7、図14ないし図18に基づいて説明すれば、以下の通りである。なお、説明の便宜上、前記の実施の形態の図面に示した部材と同一の機能を有する部材には、同一の符号を付記してその説明を省略する。

【0065】本実施の形態における機械翻訳処理は、図1に示す実施の形態1とほぼ同様の構成を有しているが、ファイル名判定のための規則が図16に示すように定められていて、図1の記憶手段11に記憶されている。ここでは、2種類の条件を図示している。すなわち、1つ目は、ファイル名の末尾が「.html」であるという条件である。2つ目は、ファイル名の末尾が「.htm」であるという条件である。これらの条件のうちのどれか一つが適合すれば、前編集記号は無効にされる。図中、IDとは、前編集記号処理を行わない条件の番号を示すものである。また、ファイル管理処理手段12の働きで、記憶手段11はファイルシステムとしても扱われ、記憶手段11内には図17に示すようなファイル管理情報が記録された領域が存在する。すなわち、ファイル名、図示しないディスク等内のこのファイルの記録された物理的位置を示す開始位置、およびファイルの大きさ(データ量)を示すデータ長を示している。開始位置として、シリンダ、トラック、セクタを示している。

【0066】本実施の形態による装置の動作の過程は、図18と図6のフローチャートに示すようになってい。まず図18に示すように動作し、図18に示す過程を終了したら図6に示すように動作する。

【0067】まず、図18のステップS201で、フラグ変数peのバッファ領域がフラグバッファ13hに確保され、フラグ変数peに値0が格納される。次にS202で、ファイル入力処理手段2cによって、図7の原文書が記憶手段11から読み込まれ、原文書バッファ1

3aに格納される。ここで、原文書が記憶手段11から読み込まれる場合は、ファイルシステムからの入力という扱いになるので、一旦、図17に示すようなファイル管理情報の領域がアクセスされ、ここで、読み込む文書のファイル名・データ長・開始位置等の情報を取得して、この情報に基づいて目的の文書が読み込まれる。このときに得られたファイル名が前編集処理しない条件に合致するか否かが、S203で、ファイル名判定処理手段8によって、図16のファイル名判定規則を参照することにより判定される。いま、読み込む文書のファイル名が「sharp.html」であるとする、このファイル名の末尾は「.html」なので、図16のファイル名判定規則に従って、この文書は前編集処理しない条件に合致しているとされ、フラグ変数peは0のまま1000に進む。

【0068】もし、ファイル名が前編集処理しない条件に合致しないとファイル名判定処理手段8によって判定されたら、S203からS204に進んで、フラグ変数peに値1が格納され、そして1000に進む。

【0069】以下の動作は図6のフローチャートに沿ったものとなり、実施の形態1による機械翻訳処理の動作と同様のものとなる。よって、原文書が図7に示すものである場合、フラグ変数peが0であるなら翻訳結果は図14のようになり、フラグ変数peが1であるなら翻訳結果は図15のようになる。

【0070】〔実施の形態3〕本発明のさらに他の実施の形態について図1、図6、図7、図14、図15および図19に基づいて説明すれば、以下の通りである。なお、説明の便宜上、前記の実施の形態の図面に示した部材と同一の機能を有する部材には同一の符号を付記してその説明を省略する。

【0071】本実施の形態における機械翻訳処理は、図1に示す実施の形態1とほぼ同様の構成を有しているが、本実施の形態による機械翻訳処理の動作の過程は、図19と図6のフローチャートに示すようになっていく。まず図19に示すように動作し、図19に示す過程を終了したら図6に示すように動作する。

【0072】まず、図19のステップS301で、フラグ変数peのバッファ領域がフラグバッファ13hに確保され、フラグ変数peに値0が格納される。次に、S302で、利用者はハンド入力手段2aを通じて、原文書の入力とその入力方法を指示する。次にS303で、前編集記号処理可否決定処理手段10によって、入力方法による動作の分岐が行われる。今、入力方法としてネットワークからの入力が指示されたとする、S308に進み、フラグ変数peは0のままである。そして、ネットワーク対応入力手段2bによって、図7の原文書がネットワークから読み込まれて原文書バッファ13aに格納され、1000に進む。

【0073】もし、入力方法としてネットワークからの

入力以外の方法が指示された場合、S303からS304またはS306に進んで、フラグ変数peに値1が格納され、それぞれハンド入力手段またはファイルシステムから原文書が入力され（S305、S307）、1000に進む。

【0074】以下の動作は図6のフローチャートに沿ったものとなり、実施の形態1による機械翻訳処理による動作と同様のものとなる。よって、原文書が図7に示すものである場合、フラグ変数peが0であるなら翻訳結果は図14のようになり、フラグ変数peが1であるなら翻訳結果は図15のようになる。

【0075】

【発明の効果】以上のように、本発明の請求項1の機械翻訳装置は、ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を作成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、文書形式判定規則データを記憶している記憶手段と、上記原文書の文書形式が上記文書形式判定規則データに合致するか否かを判定する文書形式判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を作成するか従わずに訳文書を作成するかを決定する前編集記号処理可否決定手段とを備えている構成であり、請求項2記載の翻訳プログラムを記録した媒体は、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、記憶手段に文書形式判定規則データを記憶させ、制御手段に、原文書の文書形式が文書形式判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を作成するか、前編集記号に従わずに訳文書を作成するかを決定させるものである。

【0076】従って、文書形式の判定によって、前編集記号と誤認する可能性のある記号類を含む形式の文書であるかどうか判定され、その結果に応じて前編集記号処理の可否決定を行い前編集処理が無効化されるので、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が翻訳実行に先立ち文書形式を判断する必要がなくなり、利用者の手で前編集記号を初期設定から変更あるいは無効化しておく必要がなくなる。

【0077】それゆえ、文書の翻訳において、利用者の手間を省き、操作を簡略化することができるという効果を奏する。

【0078】本発明の請求項3記載の機械翻訳装置は、ソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を作成すること

により、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、ファイル名判定規則データを記憶している記憶手段と、上記原文書のファイル名が上記ファイル名判定規則データに合致するか否かを判定するファイル名判定手段と、上記判定結果如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段とを備えている構成であり、請求項 4 記載の翻訳プログラムを記録した媒体は、記憶手段に記憶されたソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、記憶手段にファイル名判定規則データを記憶させ、制御手段に、原文書の文書形式がファイル名判定規則データに合致するか否かを判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させるものである。

【0079】従って、ファイル名の判定によって、前編集記号と誤認する可能性のある記号類を含む形式の文書であるかどうか判定され、その結果に応じて前編集記号処理の可否決定を行い前編集処理が無効化されるので、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が翻訳実行に先立ちファイル名を判断する必要がなくなり、利用者の手で前編集記号を初期設定から変更あるいは無効化しておく必要がなくなる。

【0080】それゆえ、文書の翻訳において、利用者の手間を省き、操作を簡略化することができるという効果を奏する。

【0081】本発明の請求項 5 記載の機械翻訳装置は、入力手段から入力されるソース言語で表された文書である原文書に基づき、原文書内の特定文字記号列である前編集記号に従って、ターゲット言語で表された文書である訳文書を生成することにより、ソース言語で表された文書をターゲット言語で表された文書に翻訳する機械翻訳装置において、上記入力手段の種類如何により、上記前編集記号に従って訳文書を生成するか従わずに訳文書を生成するかを決定する前編集記号処理可否決定手段を備えている構成であり、請求項 6 記載の翻訳プログラムを記録した媒体は、入力手段より入力されるソース言語で表された原文書を、制御手段によりターゲット言語で表された訳文書に翻訳させるための翻訳プログラムを記録した媒体であって、前記翻訳プログラムは、制御手段に、入力手段の種類を判定させ、判定結果如何により原文書に含まれる前編集記号に従って訳文書を生成するか、前編集記号に従わずに訳文書を生成するかを決定させるものである。

【0082】従って、原文書の入力の際に使用された入力手段の種類に基づき、前編集記号と誤認する可能性の

ある記号類を含む形式の文書であるかどうか判定され、その結果に応じて前編集記号処理の可否決定を行い前編集処理が無効化されるので、機械翻訳において、前編集記号で指定された処理を行うか否かを、装置の利用者が翻訳実行に先立ち原文書の入力の際に使用された入力手段の種類を判断する必要がなくなり、利用者の手で前編集記号を初期設定から変更あるいは無効化しておく必要がなくなる。

【0083】それゆえ、文書の翻訳において、利用者の手間を省き、操作を簡略化することができるという効果を奏する。

【図面の簡単な説明】

【図 1】本発明に係る機械翻訳装置の一構成例を示す図である。

【図 2】機械翻訳における翻訳レベルを示す説明図である。

【図 3】前編集記号とその機能を記述したデータの例を示す説明図である。

【図 4】文書形式判定規則の例を示す説明図である。

【図 5】原文入力処理を示すフローチャートである。

【図 6】原文入力以降の翻訳処理を示すフローチャートである。

【図 7】原文書の例を示す説明図である。

【図 8】翻訳処理中の原文バッファの内容の例を示す説明図である。

【図 9】翻訳処理中の訳文バッファの内容の例を示す説明図である。

【図 10】翻訳処理中の原文バッファの内容の例を示す説明図である。

【図 11】翻訳処理中の訳文バッファの内容の例を示す説明図である。

【図 12】翻訳処理中の原文バッファの内容の例を示す説明図である。

【図 13】翻訳処理中の訳文バッファの内容の例を示す説明図である。

【図 14】原文書が前編集処理を無効にして翻訳された結果としての訳文書バッファの内容の例を示す説明図である。

【図 15】原文書が静編集処理を有効にして翻訳された結果としての訳文書バッファの内容の例を示す説明図である。

【図 16】本発明に係る機械翻訳装置の他の構成例におけるファイル名判定規則の例を示す説明図である。

【図 17】ファイル管理情報データの例を示す説明図である。

【図 18】原文入力処理を示すフローチャートである。

【図 19】本発明に係る機械翻訳装置のさらに他の構成例における原文入力処理を示すフローチャートである。

【図 20】一般的な機械翻訳装置の一例を示すブロック図である。

【図21】従来の前編集記号変更機能付き機械翻訳装置の一例における、利用者によって定義された前編集記号データが格納されたバッファの内容例を示すブロック図である。

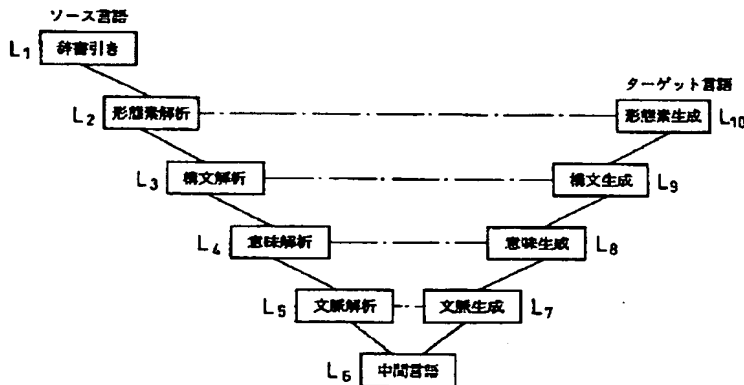
【符号の説明】

- 1 動作制御手段
- 2 a ハンド入力手段（入力手段）
- 2 b ネットワーク対応入力手段（入力手段）
- 2 c ファイル入力処理手段
- 3 出力手段
- 4 バスライン
- 5 翻訳モジュール
- 5 a 辞書引き・形態素解析処理手段
- 5 b 前編集記号処理手段
- 5 c 構文解析処理手段
- 5 d 構文変換処理手段
- 5 e ターゲット言語生成処理手段

- * 6 文切り出し処理手段
- 7 訳文書生成処理手段
- 8 ファイル名判定処理手段
- 9 文書形式判定処理手段
- 10 前編集記号処理可否決定処理手段
- 11 記憶手段
- 12 ファイル管理処理手段
- 13 動作用メインメモリ
- 13 a 原文書バッファ
- 13 b 原文バッファ
- 13 c 辞書引きバッファ
- 13 d 変換前の構文バッファ
- 13 e 変換後の構文バッファ
- 13 f 訳文バッファ
- 13 g 訳文書バッファ
- 13 h フラグバッファ

*

【図2】



【図3】

ID	前編集記号	作用
1	n-	品詞指定（名詞）
2	v-	品詞指定（動詞）
3	a-	品詞指定（形容詞）
4	//	文分割
5	{	無制限箇所開始
6	}	無制限箇所終了

前編集記号の例

【図4】

ID	前編集記号処理しない条件
1	文書中に "http://" とする文字列が含まれている
2	文書内容が "<html>" で始まって "</html>" で終わっている

文書形式判定規則の例

【図7】

```
<html>
<body>
The URL is http://www.sharp.co.jp/ .
</body>
</html>
```

原文書の例

【図9】

URLは http://www.sharp.co.jp/ である。

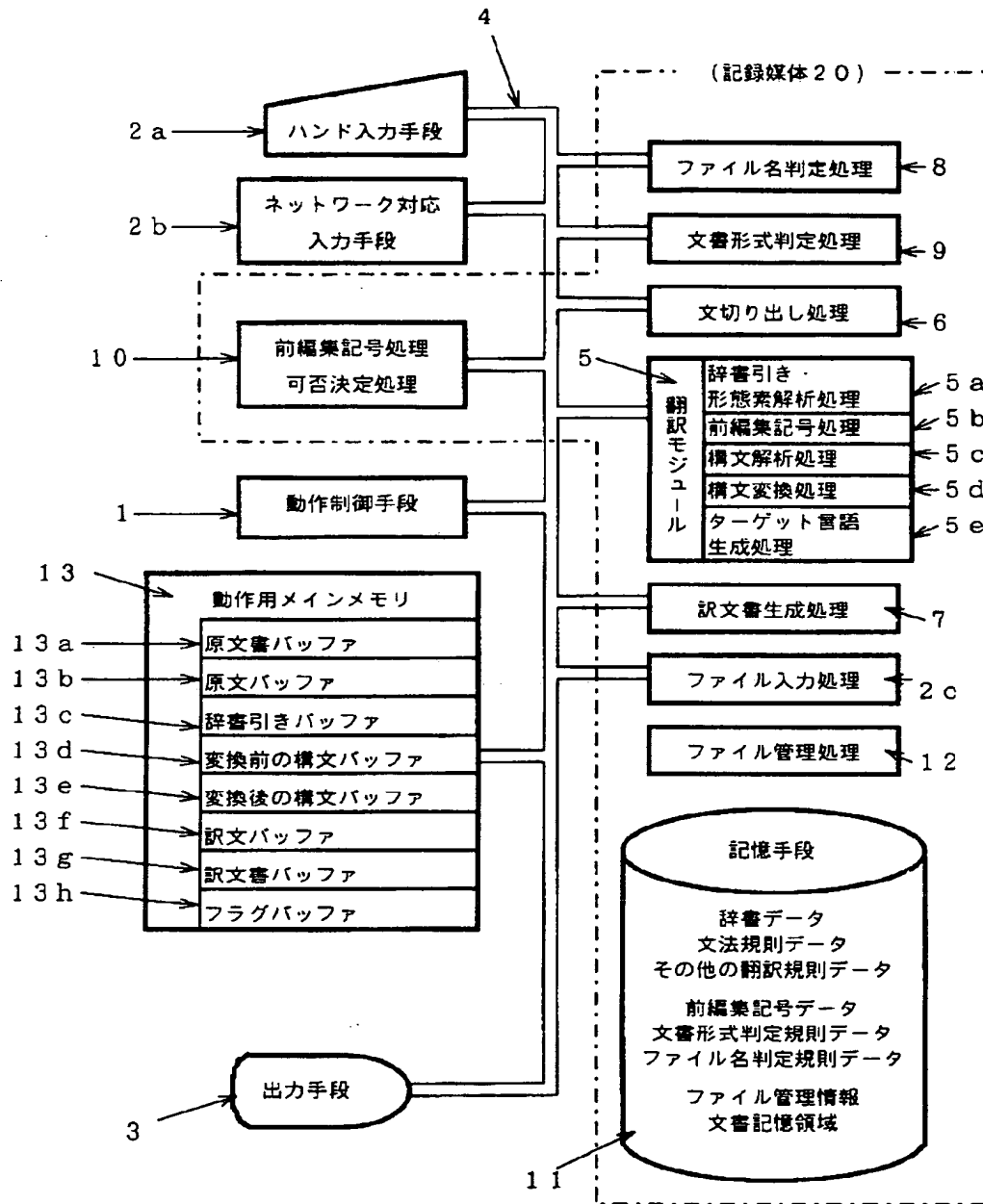
訳文バッファ

【図8】

The URL is http://www.sharp.co.jp/ .

原文バッファ

【図1】



【図10】

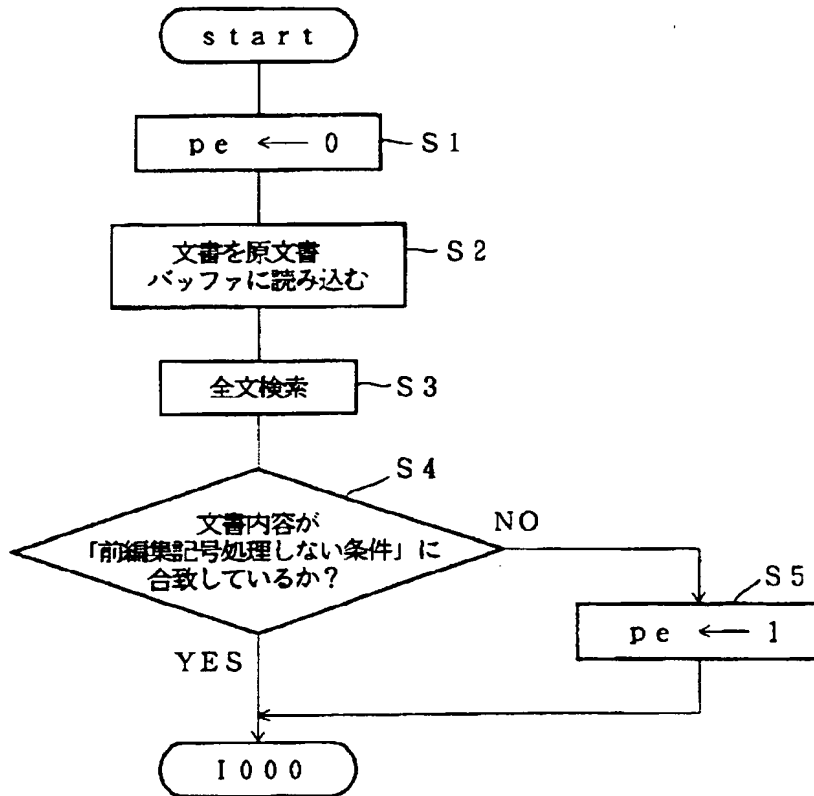
【図11】

【図12】

【図13】

The URL is http:	URLは、http: である	www. sharp. co. jp/ .	www. sharp. co. jp/ .
前編集処理後の原文バッファ	訳文バッファ	前編集処理後の原文バッファ	訳文バッファ

【図5】



【図14】

```

<html>
<body>
URLは、http://www.sharp.co.jp/ である。
</body>
</html>
  
```

訳文書バッファ

【図15】

```

<html>
<body>
URLは、http: である www.sharp.co.jp/。
</body>
</html>
  
```

訳文書バッファ

【図16】

ID	前編集記号処理しない条件
1	ファイル名の末尾が ".html" である
2	ファイル名の末尾が ".htm" である

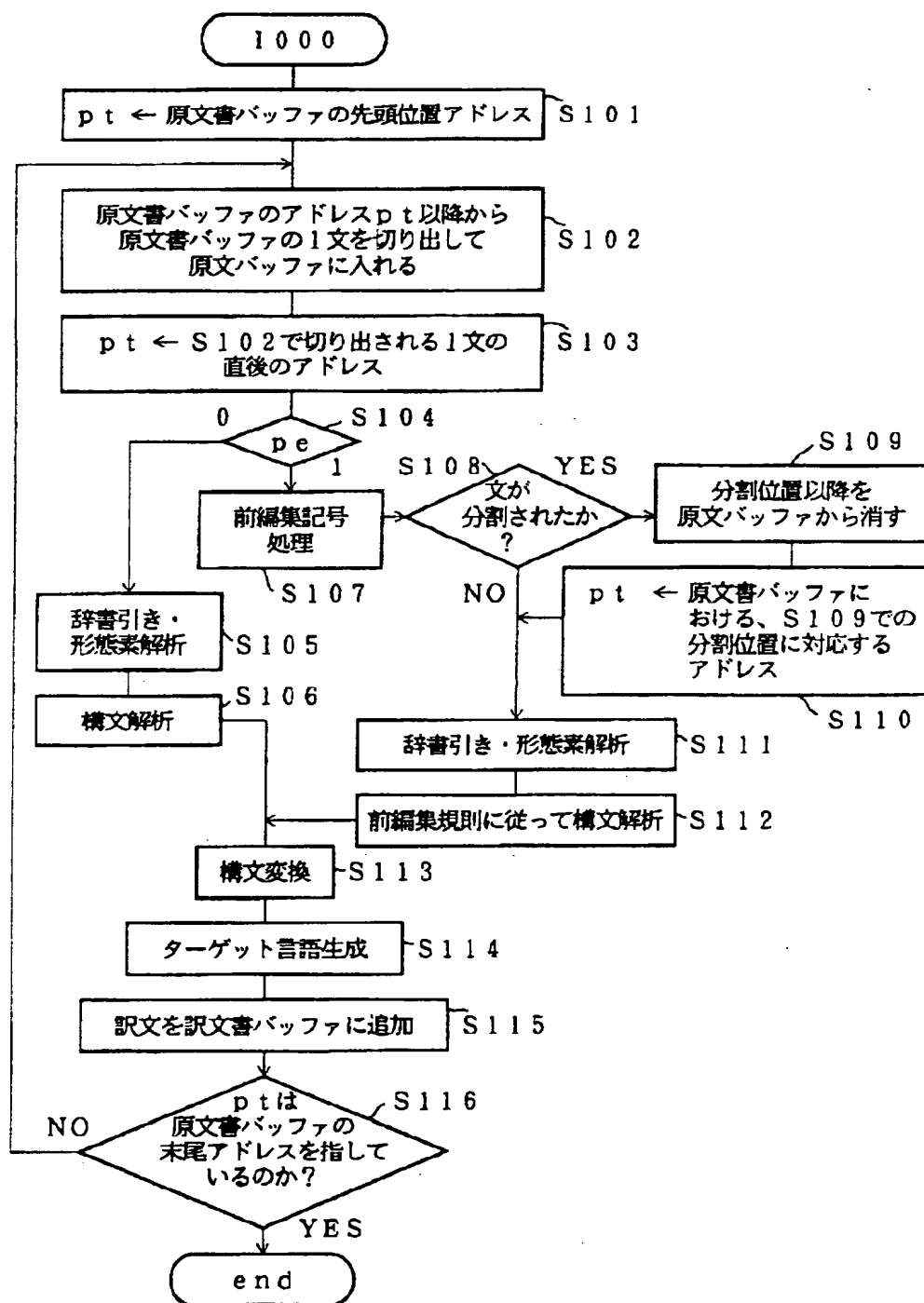
ファイル名判定規則の例

【図17】

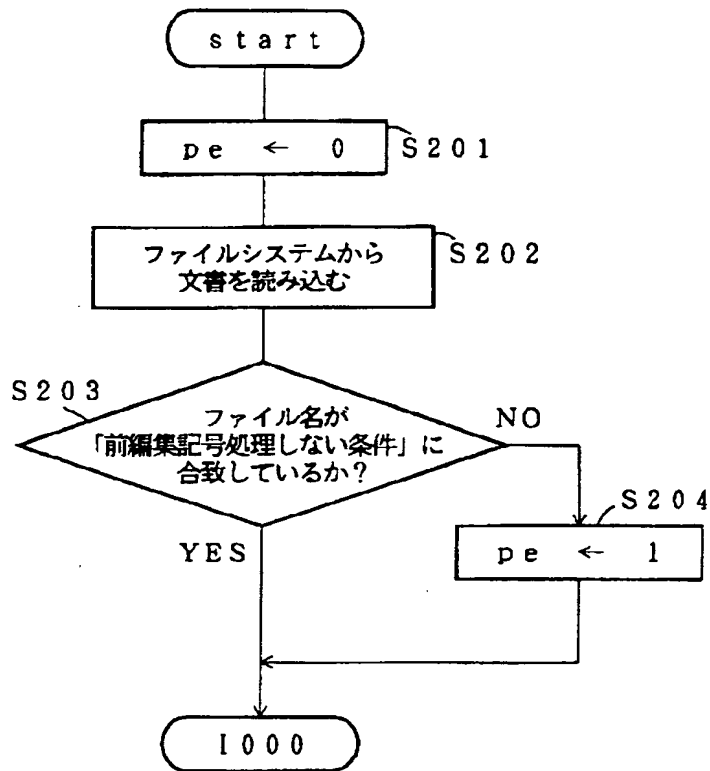
ファイル名	開始位置			データ長
	シリンダ	トラック	セクタ	
sharp.html	26	3	2	67

ファイル管理情報

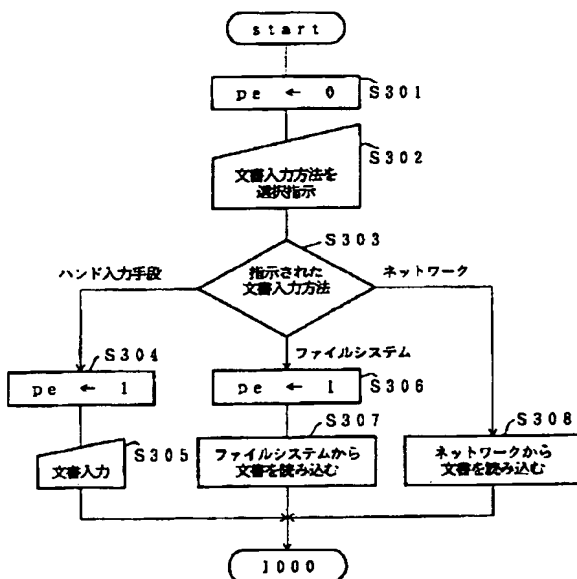
【図6】



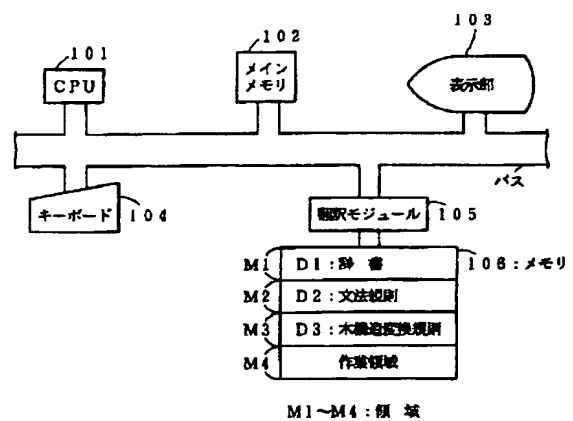
【図18】



【図19】



【図20】



【図21】

D:バッファ		K0:既定記号 ID:識別記号				
ID	K0	意味:MA	利用者定義 フラグ:F	利用者定義 記号:K1	利用者定義 記号:K2	利用者定義 記号:K3
1	n_	名詞	1	noun_		
2	v_	動詞	0			
3	a_	形容詞	0			
4	d_	副詞	0			
5	v1_	第1文型	0			
6	v2_	第2文型	0			
7	v3_	第3文型	0			
8	v4_	第4文型	0			
9	v5_	第5文型	0			
10	uj_	ユーザ辞書	0			
11	sj_	専門用語辞書	0			
12	kj_	基本語辞書	0			
13	<<	フレーズ指定開始記号	0			
14	>>	フレーズ指定終了記号	0			
15	{	部分語彙開始記号	0			
16	}	部分語彙終了記号	0			
17	{	非対象部分語彙記号	0			
18	}	非対象部分語彙終了記号	0			
19	{	無関係指定開始記号	0			
20	}	無関係指定終了記号	0			
21	//	分割語彙指定記号	1	:	,	